ToxCast Embryonic Stem Cell (H9) Profiling for Predicting Developmental Toxicity

Thomas B. Knudsen[1] Todd J. Zurlinden[1], Katerine S. Saili[1], and Nancy C. Baker[1,2]
U.S. Environmental Protection Agency, Office of Research and Development
[1]National Center for Computational Toxicology and [2]Leidos – Research Triangle Park, NC 27711
knudsen.thomas@epa.gov | 919-541-9776

Introduction

The protocol commonly used to test for prenatal developmental toxicity (i.e., OECD TG 414) is based on observation of fetal malformations, usually in pregnant rats and/or rabbits. EPA is evaluating new approach methodologies (NAMs) that can be used to quickly evaluate the human toxicity potential of chemicals with less reliance on animal testing.

ToxCast generates *in vitro* data on thousands of chemicals utilizing high-throughput screening (HTS) methods. To increase the assay space for predicting human developmental toxicity, we profiled ToxCast chemicals in the devTOX^{qP} assay [1]. The assay measures a critical drop in the ornithine/cystine ratio H9 hESC culture medium. Data for 1062 chemicals was pipelined in ToxCast to generate the STM dataset (pending release); a positive signal for developmental toxicity was elicited by 183 (17%) [2].

Here, we assessed the predictive value of the STM data with a subset of 432 ToxCast chemicals tested for prenatal developmental toxicity.

[1] Palmer JA, Smith AM, Egnash LA, Conard KR, West PR, Burrier RE, Donley ELR and Kirchner FR (2013) Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. BDRB 98: 343-363.

[2] Zurlinden TJ, Saili KS, Rush N, Kothiya P, Judson RS, Houck KA, Hunter ES, Baker NC, Palmer JA, Thomas RS and Knudsen TB (2019) Profiling the ToxCast library with a pluripotent human (H9) embryonic stem cell assay for developmental toxicity (manuscript cleared for submission).

Specific Aims

- Binary classification to correlate STM responses with animal models of human prenatal developmental toxicity captured from ToxRefDB and other sources of *in vivo* data; and
- Recursive partitioning: mining *in vitro* bioactivity profiles of >800 ToxCast *in vitro* assays to assess the value of the STM dataset for predicting human developmental toxicity.

1. Binary classification

Condition	BM-42	Base	Low	Medium	High
TP	17	85	60	35	19
FP	0	14	37	23	9
FN	9	217	127	51	11
TN	16	116	208	176	88
n	42	432	432	285	127
sensitivity	0.654	0.281	0.321	0.407	0.633
specificity	1.000	0.892	0.849	0.884	0.907
Rand ACC	78.6%	46.5%	62.0%	74.0%	84.3%
PPV	1.000	0.859	0.619	0.603	0.679
NPV	0.640	0.348	0.621	0.775	0.889
BAC	82.0%	60.3%	62.0%	68.9%	78.4%
MCC	0.647	0.190	0.202	0.332	0.554

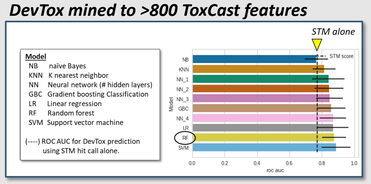
BM-42 *in vivo* benchmark of 42 ToxCast compounds identified from literature
Base adds ToxRefDB chemicals with a developmental lowest effect level (dLEL)
Low classifies ToxRefDB positives with dLEL ≤ 200 mg/kg/day in rat OR rabbit
Medium dLEL ≤ 200 mg/kg/day and < mLEL (maternal LEL) in rat OR rabbit
High concordant for positive and negative (no dLEL ≥ 1000 mg/kg/day)

in vivo / *in vitro*		
	TP	FP
	FN	TN

TP true positive, FP false positive, FN false negative, TN true negative
Regular accuracy (Rand ACC) based on sensitivity and specificity
Balanced accuracy (BAC) - positive, negative predictive value (PPV, NPV)
Mathews Correlation Coefficient (MCC) indicates level of confidence

2. Recursive analysis (*high-stringency model*)

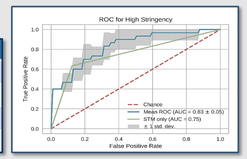
DevTox mined to >800 ToxCast features



- Machine-learning with 5-fold cross-validation on train/test split;
- ~200 ToxCast features correlated with developmental toxicity;
- STM was the top-weighted feature but others augment predictivity.

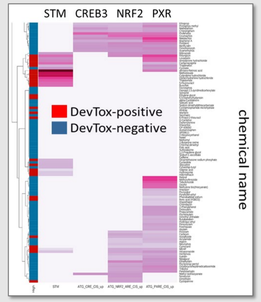
Top features following recursive elimination by RF

Feature	*Assay read-out (what the feature measures)*
STM_ORN/CYSS_dn	critical effect of the hESC biomarker
ATG_CRE_cis_up	cis-acting reporter activation via cAMP/CREB
ATG_NRF2_ARE_cis_up	NFE2L2 antioxidant response element
ATG_PXR_cis_up	cis-acting reporter activation via PXR/PXRE



- 4 ToxCast features tied for 1st after removing those unimportant to predictive model;
- area under the ROC (AUC) compares STM-only with STM-augmented models;
- slight boost in performance evident based on z-score filtering at 1 standard deviation.

3. Biological plausibility



- Cluster analysis of STM (hESCs) and CREB/NRF2/PXR (Hep2G) hit-calls show a generally inverted correlation (n=127 chemicals in the high-stringency model);
- assuming STM measures an adverse response, the inverted correlation infers an adaptive response for the following pathways:
  CREB →(+) unfolded protein response (*UPR pathway*)
  NRF2 →(+) cellular redox control (*ARE pathway*)
  PXR →(+) xenobiotic metabolism and export (*XME pathway*).
- sets up a hierarchical rules-based workflow for building a decision workflow:
  Rule 1: STM(+) & CREB3(-) predicts TP condition in 19 of 22 cases (86.4%)
  Rule 2: CREB3/NRF2/PXR (+) predicts TN condition and overrides STM(+) call
  Rule 3: STM(-) & PXR(+) OR NRF2(+) predicts TN condition in 63 of 69 cases (91.3%)
  Rule 4: STM(-) & CREB3/NRF2/PXR(-) condition predicts TN in 25 of 30 cases (83.3%)

4. Binary classification refined *(work in progress)*

Refined binary classification (based on Rules 1 – 4)

Condition	Low	Low+	Medium	Medium+	High	High+
TP	60	50	35	33	19	19
FP	37	24	23	13	9	3
FN	127	137	51	53	11	11
TN	207	220	175	185	88	94
n	431	431	284	284	127	127
sensitivity	0.321	0.267	0.407	0.618	0.633	0.633
specificity	0.848	0.902	0.884	0.976	0.907	0.969
Rand ACC	61.9%	62.6%	73.9%	89.8%	84.3%	89.0%
PPV	0.619	0.676	0.603	0.875	0.679	0.864
NPV	0.620	0.616	0.774	0.902	0.889	0.895
BAC	61.9%	64.6%	68.9%	88.9%	78.4%	87.9%
MCC	0.201	0.222	0.331	0.679	0.554	0.676

↑ stringency →

- Recalculation of PPV and NPV for the low, medium, and high-stringency performance models based on augmentation with the STM/CREB3/NRF2/PXR data (denoted by +);
- Balanced accuracy reached 90% in the augmented model v1; studies are in progress to further bolster the model with other ToxCast assays.

5. Conclusions

The STM biomarker surfaced as the top weighted feature for developmental toxicity in the ToxCast *in vitro* portfolio. Balanced accuracy (BAC) reached 78% with well-supported evidence for developmental toxicity but depreciated to 62% with when criteria for developmental toxicity was relaxed. Augmenting the STM response with data from several ToxCast assays that measure adaptive cellular responses (UPR, ARE, XME pathways) improved positive predictive value to BAC ~90%.